# Audience Silhouettes: Peripheral Awareness of Synchronous Audience Kinesics for Social Television

**Radu-Daniel Vatavu**
University Stefan cel Mare of Suceava
Suceava 720229, Romania
vatavu@eed.usv.ro

## ABSTRACT

We introduce *audience silhouettes* for TV, which are visual representations of viewers' body movements displayed in real-time on top of television content. With their minimal visual cues and their ability to convey presence and to leverage interactions via non-verbal *kinesics*, audience silhouettes are strong candidates for implementing Oehlberg *et al.*'s theater metaphor of an unobtrusive social TV system [37]. In a user study, we found our participants connecting well to the on-screen silhouettes, while their television watching experience was perceived more enjoyable. We also report viewers' body movement behavior in the presence of on-screen silhouettes, which we characterize numerically with new measures (*e.g.*, average body movement) and we report experimental findings; *e.g.*, we found that the number of silhouettes influences viewers' body movements and the body postures they adopt and that women produce more body movement than men.

### Author Keywords

Television; Kinect; audience silhouettes; social TV; motion capture; whole body gestures; peripheral awareness; user study; user experience; SUS; augmented TV; kinesics.

### ACM Classification Keywords

H.5.1 Multimedia Information Systems: Artificial, augmented, and virtual realities; H.5.2. User Interfaces: Input devices and strategies (*e.g.*, mouse, touchscreen).

## INTRODUCTION

Social television watching at distance is supported today by a variety of smart devices and social networks. Many studies have revealed viewers' desire to feel connected, either to family and friends or to a larger community interested in the same TV shows [4,17,39]. However, despite numerous research on leveraging audience interactions with text and audio chat during synchronous television watching [19,25,28], little work has addressed non-verbal body communication, *i.e.*, kinesics, for iTV. Body movement can provide rich informational cues about one's intents or emotions [7,32,34], but today's social iTV systems offer no availability for transmitting these cues, other than through fully-disclosing video or artificial avatars with little expressive resources [14,36]. Moreover, this challenge has existed in the research agenda of social television for quite a while, *i.e.*, delivering *"presence awareness that aids communication flow"* [11] (p. 8), and Oehlberg *et al.* [37] imagined the *theater metaphor* to depict an unobtrusive social TV system, yet to be implemented at its full potential. In this work, we make one step further toward Oehlberg *et al.*'s vision by introducing *audience silhouettes*, which are body profiles of remote viewers watching the same show; see Figure 1. Audience silhouettes do not disclose a person's fine traits, such as face or clothes (as video does), but instead they communicate presence and expressive body movements in real-time.

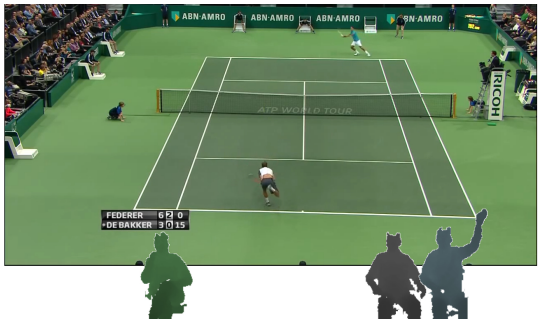


**Figure 1. Audience silhouettes are body profiles of remote viewers displayed on top of TV content. Silhouettes do not disclose a person's fine traits, such as face or clothes (as video does), but instead they communicate presence and expressive body movement with minimal visual cues; *e.g.*, color gradients assist the perception of depth and body movement.**

Our contributions are as follows: (1) we introduce the concept of *audience silhouettes* to support real-time non-verbal communication during social television watching; (2) we evaluate viewers' perceptions of audience silhouettes, and we report findings on the *peripheral awareness* of kinesics; *e.g.*, our participants felt connected to silhouettes, which made the TV watching experience more enjoyable; (3) we introduce new measures to characterize viewers' body movements in relation to audience silhouettes; *e.g.*, we found that the number of silhouettes affects body movement and adopted body postures; and (4) we introduce Motion-Amount Images to visualize and interpret body movements. As we have barely scratched the potential of real-time kinesics for social iTV in this work, we hope that this first exploration of audience silhouettes will inspire the community to explore kinesics further for designing enriched experiences for social interactive television.


*TVX'15*, June 3–5, 2015, Brussels, Belgium.

ACM 978-1-4503-3526-3/15/06...$15.00
http://dx.doi.org/10.1145/2745197.2745207

## RELATED WORK

In this section we discuss related work on designing systems to enrich user experience during television watching, we review prior work on designing for social television, and we connect to existing research in kinesics and body gestures.

### Social television watching at distance

The social aspect of watching television by viewers that are geographically located at distance has been supported with communications technology. Over time, television-mediated interpersonal communication has employed many modalities, such as text, audio, and voice chats, emoticons and avatars, as well as combinations of these [14,19,26,36,40]. For example, Amigo-TV [14] is a system that combines broadcast television with communication between viewers implemented by transmission of speech, text, and emoticons to leverage a rich social experience during television watching; in Amigo-TV, viewers are represented by avatars. Nathan *et al.* [36] developed CollaboraTV, a system that supports both synchronous and asynchronous viewers under a unified interface; viewers are represented as iconic avatars forming a virtual audience at the bottom of the TV screen that can share text messages with localized speech bubbles and perform small animations (*e.g.*, avatars may raise their arms and produce the thumbs-up or thumbs-down gestures, or may turn around toward the viewer and show a happy emoticon face). Social TV and Social TV 2 [25,26] are interactive TV systems that display the watching status of remote viewers and allow sending lightweight text messages between viewers.

Text and audio have been widely researched for social iTV with findings that depend on the audience and usage context. For instance, in a controlled lab experiment involving 17 subjects, Geerts [19] found that voice chat was considered more natural and direct than text chat, but text was preferred by young viewers who had previously used it on computers. In an *in-situ* study with 5 male subjects, Huang *et al.* [28] found that their participants overwhelmingly preferred text to voice chat, but also that they often employed the system to communicate about topics not related to television content. Whatever the modality of mediating interactions, the challenge has always been to support viewers engaging in communication, while not obstructing television watching [11,37].

The iTV community has also looked at ways to better define, understand, and analyze the social dimensions of television watching in order to inform improved designs of social iTV systems and, consequently, deliver enriched user experience. For instance, Chorianopoulos [13] introduced *presence* and *type of communication* as two dimensions for analyzing the social aspects of television; in terms of presence, viewers can be collocated or at distance, while communication can be either synchronous or asynchronous. Oehlberg *et al.* [37] proposed design strategies for social interaction between distant viewers during television watching to prevent disruption of TV flow and to support fluent conversation between viewers, such as minimize disruptions in following content on the TV, isolate side conversations, and avoid drawing viewers' attention away from the TV screen. Cesar, Chorianopoulos, and Jensen [11] provided a relevant overview of the social and interaction aspects of television. Geerts and De Grooff [21] remarked the lack of sociability heuristics for evaluating social TV systems, and introduced twelve guidelines to assist practitioners in this direction, such as allowing for both synchronous and asynchronous communication, exploiting viewing behavior to engage other viewers, guaranteeing personal and group privacy, and letting users share content in a flexible manner.

Recently, social television watching has been leveraged and boosted by secondary screen applications and by a variety of social networks. These instruments allow people to interact with each other using their mobile devices, either directly, *e.g.*, by means of live chat, or indirectly by posting and following posts about TV content on various social websites. While reviewing previous work on secondary screens, Cesar, Bulterman, and Jansen [10] identified four main motivations for their usage in an interactive television environment, which are control, enrich, share, and transfer of television content. In a different study, Courtois and D'heer showed that participants mostly used their tablet devices during television watching for social networking and content search [15].

### Kinesics and body gestures

Kinesics represents the interpretation of non-verbal communication expressed with body movement, gestures, and facial expressions. While coining the term, Birdwhistell [7] also considered pre-kinesics (*i.e.*, the physiology of kinesics), micro-kinesics (*i.e.*, the study of *kines*, which represent particles of abstractable body movement), and social kinesics (*i.e.*, the use of body communication in social interaction). A large body of literature in psycholinguistics has shown that body gestures are deeply connected with language, speech, and thought [18,34], making them remarkable conveyors of information to listeners [32]. Moreover, gestures were found to facilitate the smoothness of interaction and to increase linking between interaction partners, *i.e.*, *the chameleon effect* [12], and to communicate attitudes and emotions both voluntarily and involuntarily [23]. Kleinsmith and Bianchi-Berthouze [33] compiled a survey of the literature concerned with the perception of affective body expression. Bianchi-Berthouze [6] investigated players' body engagement during whole-body gesture-controlled video games, for which she proposed a taxonomy of body movements and worked with a model describing the relationship between movement and type of engagement.

### Designing augmented TV experiences with Kinect

We employ in this work the Microsoft Kinect depth sensor to capture audiences' body movements and to display them on top of television content. Our use of depth sensing and human motion capture technology for interactive TV applications is not new in the iTV community. In fact, researchers have employed Kinect to create novel interfaces for TV and home entertainment, such as controlling the TV set with the bare palm [16], entering text on TV [38], projecting video gaming content in the periphery of the TV set [30] and in the entire room [29]. Also, researchers have used Kinect to collect viewers' preferences for TV gestures [35,42,43]. In this work, we follow this practice and employ the Kinect sensor to capture viewers' silhouettes in their environment to deliver new, enriched social experiences for television.

## PROTOTYPE

We implemented a prototype application that captures viewers' body silhouettes from the depth stream delivered by the Microsoft Kinect sensor[1], and displays them on top of television content. Other silhouettes, captured by distinct Kinect sensors installed in other locations are synchronously displayed on top of the same content (at about 13.5 fps for 3 active silhouettes, see the Results section). Silhouettes are not disclosing any traits of their viewers, such as face or clothes, as they are displayed using color gradients with darker colors showing more depth, see Figure 1 on the first page. The prototype was implemented in .NET 4.5 C# using the Microsoft Kinect SDK that distinctly marks body pixels inside depth frames, which makes user segmentation easy. In our implementation, we have followed Oehlberg *et al.*'s "Mystery Science Theater 3000" metaphor of an unobtrusive social TV system [37], for which the interface is a row of theater seats at the bottom of the TV screen. According to Oehlberg *et al.*, such an interface would be less distracting than displaying full-bandwidth video of each viewer, but it would not convey as many social cues as video. Our prototype represents the reasonable compromise for conveying kinesics of remote viewers, while minimizing the resolution for human body representation and, consequently, minimizing bandwidth for video transmission.

## USER STUDY

We conducted a study to understand the opportunity of audience silhouettes to enrich viewers' television watching experience as well as to collect user feedback in terms of perceived kinesics and overall body movement reaction and responsiveness with respect to the audience silhouettes concept.

### Participants

Fifteen (15) participants volunteered for the study (7 were females) with ages between 19 and 28 years old (average age 22.4 years, SD=2.1 years). Ten (10) participants had a technical background, while the rest were non-technical. We made sure that our participants' age range (19−28 years) was reflective of today's owners of smart TV products. For instance, "The Connected Consumer Survey 2013: TV and video" of Analysis Manson Limited reports that people in the 18−34 age group are most likely to own a smart TV and also to make full use of it, *e.g.*, to actually connect it to the Internet [1] (p. 31). Our participants reported watching TV content (either on the TV set or on some other device) for an average of 2.6 hours per day (SD=1.2 hours).

### Apparatus

The audience silhouettes application prototype ran on a 3.2GHz Quad-Core PC with Windows 7, which was connected to the TV set (Sony BRAVIA, 40 inch/102 cm diagonal) running at full HD resolution 1920×1080 pixels. The Microsoft Kinect sensor was placed on top of the TV set, and depth video frames were captured at a resolution of 320×240 pixels, while the frame rate depended on the actual CPU load (with an average 13.5 frames per second). Participants' body silhouettes were recorded as binary files, with one file per each experimental condition (see next) and 180 files in total.

[1] http://www.microsoft.com/en-us/kinectforwindows/

### Design

Our study was a within-subjects design with two factors:

1. AUDIENCE-TYPE, nominal variable having 4 conditions: NO-AUDIENCE (*i.e.*, regular television watching with no silhouette feedback), SELF-AUDIENCE (*i.e.*, only the viewer's body silhouette is displayed on screen), SINGLE-AUDIENCE (*i.e.*, the viewer's body silhouette and one audience silhouette are shown), and MULTIPLE-AUDIENCE (*i.e.*, two different silhouettes are shown on the TV screen next to the viewer's own silhouette). Figure 2 shows these conditions.
2. GENRE, nominal variable with 3 conditions: NEWSCAST, MOVIE, and SPORTS. Genres were informed by prior research that investigated the types of content that make people talk the most while watching TV and, consequently, are most suited for synchronous social iTV applications [20].

In our analysis, we also report and discuss results based on participants' GENDER, nominal factor with 2 conditions.

### Task

Participants sat in a comfortable armchair at a distance of approximately 2 meters from the TV set. Each AUDIENCE-TYPE condition was presented to each participant for 3 minutes, with a total of 12 minutes of television watching per participant. We considered that this amount of time would be sufficient to capture participants' body movement behavior, knowing that prior investigations of people watching TV showed that engaged looks generally take between 6 and 15 seconds and that staring installs after 15 seconds of continuous TV watching [27]. The order of conditions was randomized across participants. Each condition showed a sequence from a larger video file with NO-, SELF-, SINGLE-, and MULTIPLE-AUDIENCE silhouettes displayed on top. Participants were told that friends of theirs were in a different room watching the same transmission in order to create the impression of a live audience. Instead, we played recordings of previously captured audience silhouettes in order to assure the same visual stimuli for each participant (*i.e.*, the silhouettes we displayed always behaved in the same way at exactly the same time of the experiment). From this perspective, our experimental setup is actually a Wizard-of-Oz design [31]. The experimenter left the room while participants watched television to not influence their body behavior. Then, the experimenter returned and asked participants to fill a questionnaire, which is described in detail in the next section. The experiment took about 25 minutes per participant.

### Measures

We employ both *objective* and *subjective* measures to characterize and analyze participants' behavior in relation to on-screen audience silhouettes and their displayed kinesics.

The objective measures are computed from the body movement data collected with the Microsoft Kinect sensor. Specifically, the Kinect sensor delivers depth frames of the scene that we recorded at a resolution of $320 \times 240$ pixels and a maximum frame rate of 30 fps. We extracted participants' body movements from the Kinect-delivered depth scene, which we recorded as a set of body postures $P_i$, $i = 1..T$ for the entire monitored time interval $T$, with each posture $P_i$ representing a set of 3-D body points whose coordinates are expressed in

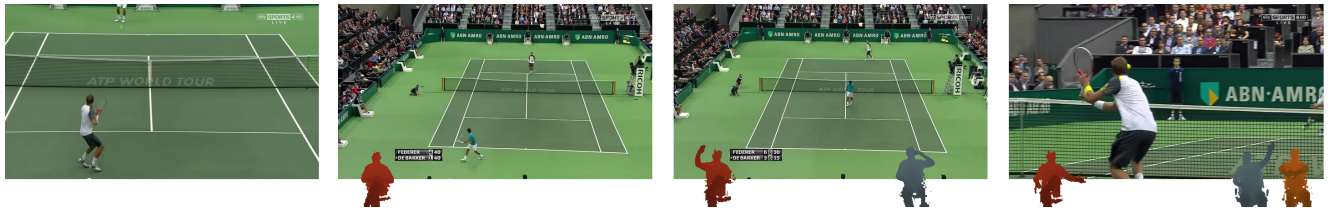

**Figure 2. The four experimental conditions employed for the AUDIENCE-TYPE factor, from left to right: NO-AUDIENCE, SELF-AUDIENCE, SINGLE-AUDIENCE, and MULTIPLE-AUDIENCE. The participant's own silhouette is displayed in red colors at the bottom-left side of the TV screen.**

meters in a system of reference centered on the Kinect sensor:

$$P_i = \{p_{i,j} = (x_j, y_j, z_j) \in \mathbb{R}^3 \mid j = 1..|P_i|\} \quad (1)$$

where $|P_i|$ is the number of points constituting posture $P_i$.

Based on this representation for body posture and movement, we define and employ the following objective measures to characterize participants' body movement behavior during television watching in the presence of audience silhouettes:

1. BODY-MOVEMENT represents the average amount of movement performed by participants during the monitored time interval defined as the average of normalized symmetric differences between time-consecutive postures $P_i$ and $P_{i+w}$:

$$\text{BODY-MOVEMENT} = \frac{w}{T} \cdot \sum_{i=1}^{T-w} \frac{|P_{i+w} \triangle P_i|}{|P_i| + |P_{i+w}|} \cdot 100\% \quad (2)$$

   where $w$ is a time window parameter for which we used 1 second (*i.e.*, $w$ averages to 13.5 fps for our dataset). The symmetric difference $\triangle$ between two point sets means that we count all the pixels that are present in $P_{i+w}$ but not in $P_i$ and vice versa, after which we normalize the result by dividing it by the total number of points subjected to comparison, *i.e.*, the cardinals of sets $P_i$ and $P_{i+w}$. We then compute BODY-MOVEMENT as the average of $\triangle$ differences between $w$-consecutive body frames. Due to this normalization process, we report BODY-MOVEMENT values as percentages, *e.g.*, an average of 12% of the participant's body pixels moved during a time duration of 60 seconds.

2. DISTINCT-POSTURES represents an indicator of the diversity of body postures produced by participants during the monitored time interval. To compute this measure, we apply the symmetric difference operator $\triangle$ (eq. 2) to all pairs of body postures for a participant in a trial and count how many postures are different by at least $\delta$=25% difference (a threshold value that we derived experimentally by visually appreciating the difference in body postures):

$$\text{DISTINCT-POSTURES} = \frac{\left|\left\{(P_i, P_j) \mid \frac{|P_i \triangle P_j|}{|P_i|+|P_j|} \geq \delta\right\}\right|}{\frac{1}{2} \cdot T \cdot (T-1)} \cdot 100\% \quad (3)$$

   where we enumerate all body posture pairs $(P_i, P_j)$, $1 \leq i < j \leq T$. Due to the normalization process, we also express DISTINCT-POSTURES values as percentages, *e.g.*, a value of 36.1% means that 36.1% of all body posture pairs of the participant in a trial were composed of postures different by at least $\delta$=25%.

3. MOVEMENT-AMPLITUDE represents the space volume in which body movement occurs:

$$\begin{aligned}\text{MOVEMENT-AMPLITUDE} = &(\max_{i=1,T} x_i - \min_{i=1,T} x_i) \cdot \\ &(\max_{i=1,T} y_i - \min_{i=1,T} y_i) \cdot \\ &(\max_{i=1,T} z_i - \min_{i=1,T} z_i)\end{aligned} \quad (4)$$

   As the Kinect sensor reports $x$, $y$, and $z$ coordinates in meters, we report MOVEMENT-AMPLITUDE in $m^3$, *e.g.*, 1.8 $m^3$ would represent the space volume in which a participant moved during a total time of say 60 seconds.

These 3 quantitative measures capture various aspects of how people move in front of the TV. For instance, while BODY-MOVEMENT reports differences in movement from frame to frame and MOVEMENT-AMPLITUDE characterizes the space in which movement takes place, DISTINCT-POSTURES measures the uniqueness and distinctiveness of one's body postures. Although more measures can be imagined to further characterize body movement, we employ in this work the minimal set capable to validate our hypotheses numerically (see next section), while we hope that readers will be inspired to try out other measures as well (see the Future Work section).

We also employ a number of 9 subjective measures collected with questionnaires, mostly as Likert scale ratings denoting the degree of participants' agreement to various statements:

1. PERCEIVED-USEFULNESS, measured on a 5-point Likert scale as evaluation of the statement *"I find the concept of TV audience silhouettes an useful one."* The 5 levels of the Likert scale are (1 to 5): strongly disagree, disagree, neither agree nor disagree, agree, and strongly agree.
2. PERCEIVED-ENJOYMENT, measured on a 5-point Likert scale as degree of agreement with the statement *"I find the concept of TV audience silhouettes an enjoyable one."*
3. PERCEIVED-DISTRACTEDNESS, measured on a 5-point Likert scale as degree of agreement with *"I find the concept of TV audience silhouettes distracting me from watching the TV program."*
4. DESIRABILITY, measured with the Microsoft Reaction Cards method[2] [5]. Participants are asked to describe the audience silhouettes concept using any of a set of 118 words, such as *appealing*, *effortless*, *impressive*, *distracting*, etc.

[2]Permission is granted to use this Tool for personal, academic and commercial purposes. If you wish to use this Tool, or the results obtained from the use of this Tool for personal or academic purposes or in your commercial application, you are required to include the following attribution: "Developed by and © 2002 Microsoft Corporation. All rights reserved".

Participants can pick as many words as they deem relevant, after which they highlight the 5 most relevant words.

5. PERCEIVED-USABILITY, measured with the System Usability Scale (SUS) tool [9]. SUS consists of 10 statements for which participants rate their degree of agreement using 5-point Likert scales, and answers are aggregated into a score ranging from 0 (low usability) to 100 (perfect score).
6. PERCEIVED-CONNECTEDNESS, measured on a 5-point Likert scale as degree of agreement with *"I felt connected with the remote person while watching television."*
7. PERCEIVED-SOCIAL-DISCOMFORT, measured on a 5-point Likert scale as degree of agreement with *"I felt discomfort seeing the remote person while watching television."*
8. PERCEIVED-SOCIAL-EXPERIENCE, measured on a 5-point Likert scale as degree of agreement with the statement *"Watching TV with another remote person that I was able to see made my watching experience more enjoyable."*
9. PERCEIVED-KINESICS, measured on a 5-point Likert scale as degree of agreement with the statement *"I was able to understand well the body language of the other person."*

### Hypotheses

We formulate the following hypotheses for our study:

- $H_1$. AUDIENCE-TYPE will influence participants' body movement behavior.
- $H_2$. The GENRE of displayed TV content will influence participants' body movement behavior.
- $H_3$. Men and women will react differently to audience silhouettes in terms of their body movement behavior.

## RESULTS

We collected a total number of 146,087 body postures from 15 participants representing 180 minutes of body movement data recorded at an average frame rate of 13.5 frames per second. In the following, we analyze participants' body movements with our quantitative measures and we look at participants' self-reported experience with TV audience silhouettes that we measured using SUS, report cards, Likert scale ratings, and comments elicited with open-ended questions.

### Participants' body movements

We found a significant effect of AUDIENCE-TYPE on participants' average BODY-MOVEMENT ($\chi^2_{(3,N=45)}$=30.408, $p$<.001), which increased from 7.9% (SD=2.8%) for the NO-AUDIENCE condition to 8.2% (SD=3.4%) for SELF-AUDIENCE, 8.8% (SD=3.8%) for SINGLE-AUDIENCE, and reached the maximum value of 9.7% (SD=3.3%) when participants were subjected to the MULTIPLE-AUDIENCE condition; see Figure 3. Follow-up post-hoc Wilcoxon signed-rank tests showed significant differences (Bonferroni corrected at $p$=.05/6=.0083) only between MULTIPLE-AUDIENCE and all the other three AUDIENCE conditions. There were no significant differences detected between body movement collected during the NO-, SELF-, and SINGLE-AUDIENCE conditions. These results suggest that more on-screen silhouettes were able to influence participants' behavior to change significantly in terms of produced body movement, which can be interpreted

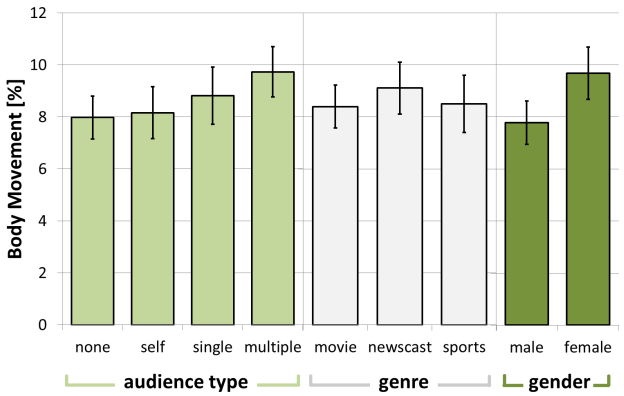


**Figure 3. Participants' average percentages of BODY-MOVEMENT computed for the AUDIENCE-TYPE, GENRE, and GENDER experimental conditions. NOTE: error bars show 95% CIs.**

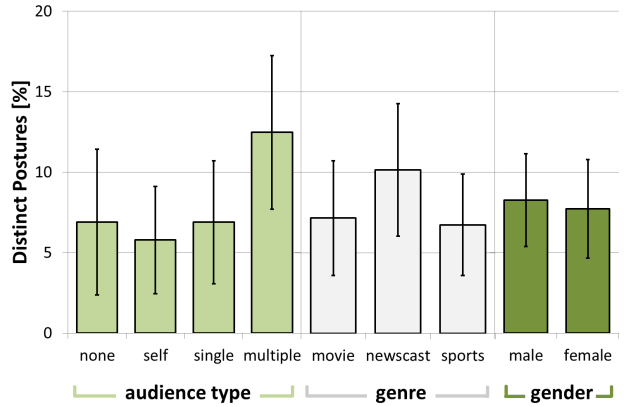


**Figure 4. Participants' average percentages of DISTINCT-POSTURES computed for the AUDIENCE-TYPE, GENRE, and GENDER experimental conditions. NOTE: error bars show 95% CIs.**

as greater involvement with the on-screen audiences. We detected no significant effect of GENRE on participants' BODY-MOVEMENT ($\chi^2_{(2,N=60)}$=5.246, $n.s.$), which shows that our selected genre content was not able to influence body behavior on its own. Note however that different results could be obtained by exposing participants to other TV content and genres that might possess different capabilities to trigger emotional response, not examined in this work. However, we found a significant effect of GENDER ($t_{(178)}$=−3.907, $p$<.001), showing that women produced significantly more body movement than men during television watching (9.7%, SD=3.4% versus 7.8%, SD=3.4%). This finding confirms for our specific television application scenario previous research results that showed women generally expressing more emotion than men in terms of their non-verbal behavior, such as smiles, laughs, head and body movement [24].

We found a significant effect of AUDIENCE-TYPE on participants' percentage of distinctly adopted body postures ($\chi^2_{(3,N=45)}$=14.217, $p$<.01). The maximum percentage of DISTINCT-POSTURES (12.5%, SD=16.4%) occurred for the MULTIPLE-AUDIENCE condition; see Figure 4. Follow-up Wilcoxon signed-rank tests showed significant differences (Bonferroni corrected at $p$=.05/6=.0083) only between the (MULTIPLE-AUDIENCE and NO-AUDIENCE) and (MULTIPLE-

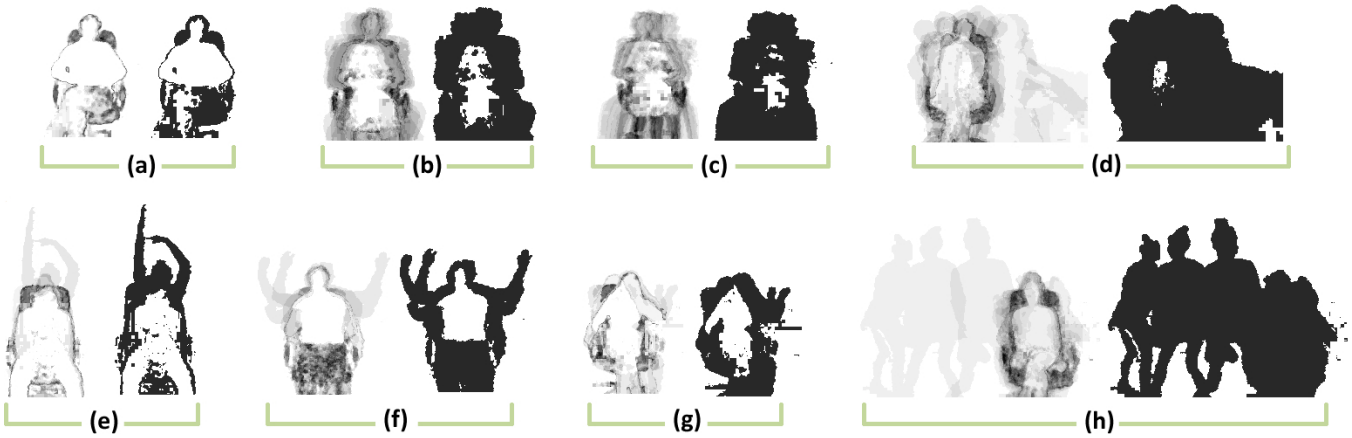


**Figure 5. Examples of Motion-Amount Images (gray levels) and Motion-Energy Images (black & white) computed from one minute of recording. Note the various body behaviors of our participants, such as sitting and crossing legs (a), leaning left and right (b), swiveling (c), reaching for an object (d), trying to attract the attention of the on-screen silhouettes (e), (f), (g), and even standing up and walking around (h).**

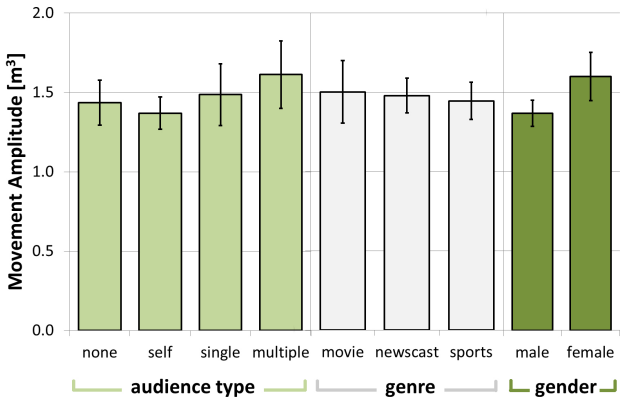


**Figure 6. Participants' average MOVEMENT-AMPLITUDE computed for the AUDIENCE-TYPE, GENRE, and GENDER experimental conditions. NOTE: error bars show 95% CIs.**

AUDIENCE and SELF-AUDIENCE) pairs of conditions. These results show again that our participants' body behavior was influenced more by the presence of more on-screen silhouettes. We detected no significant effects of GENRE on participants' DISTINCT-POSTURES ($\chi^2_{(2,N=60)}$=1.636, $n.s.$), and no significant effect of GENDER ($t_{(178)}$=0.257, $n.s.$).

We found no significant effect of AUDIENCE-TYPE nor GENRE on MOVEMENT-AMPLITUDE ($\chi^2_{(3,N=45)}$=7.664, $n.s.$, and $\chi^2_{(2,N=60)}$=0.795, $n.s.$ at $p$=.05). There was however a significant effect of GENDER ($t_{(178)}$=−2.657, $p$<.01), showing that women produced slightly (14%) more ample movements than men did (1.6 $m^3$, SD=0.7 $m^3$ versus 1.4 $m^3$, SD=0.4 $m^3$); see Figure 6.

These results validate hypotheses $H_1$ and $H_3$, but not $H_2$. To understand participants' body movements in more detail, we generated Motion-Energy Images (MEIs) and a variant of Motion-History Images (MHIs) [8] from our collected body posture data. Motion-Energy Images are black and white image representations of motion, which are computed as cumulative differences of motion occurring between consecutive video frames [8] (p. 260). Motion-History Images are gray-level images that reflect the temporal aspect of motion as it unfolds in time with brighter colors showing motion that is more recent [8] (p. 260). Instead of depicting time, we used gray levels to illustrate the *amount* of body movement and, consequently, we compute Motion-Amount Images (MAIs). Note that while MEIs show *where* body movement occurred, MAIs reveal *how much* movement occurred at each point in the captured scene. Figure 5 shows MAI and MEI images computed for some of our participants. The Appendix (Figure 9) provides all the MAI images generated for all the 180 experimental recordings (=15 participants × 4 AUDIENCE-TYPES × 3 GENRES) so that readers can form a better understanding of all our participants' body movement behavior.

Overall, we found participants producing a variety of body movements suggestive of their experience of watching and interacting with TV audience silhouettes, even if our setting was a laboratory-controlled one. (It is reasonably to assume that a larger variety of body movements will be produced in the familiarity of one's own living room, which is an investigation that we leave for future work.) For example, we observed that some participants did not bother at all interacting with the on-screen silhouettes, while they simply preferred to sit comfortably in the armchair, which is characteristic for the lean-back paradigm; see Figures 5a, 5b, and 5c. This behavior can be explained by participants actually focusing on the TV content and ignoring the audience or by deliberately self-restraining their body movements because of the unfamiliar laboratory environment. On the other hand, other participants felt more comfortable and we were able to see their willingness to interact and communicate with the on-screen audience silhouettes in a lean-forward way; see Figures 5e, 5f, and 5g for some examples of participants trying to attract the attention of the on-screen silhouettes. Indeed, when asked about their behavior vis-a-vis silhouettes, nearly all participants (14/15=93%) said they tried to interact with the on-screen silhouettes, from *a little* (12 out of 15 responses) to *a lot* (2 participants). Other body behavior included reaching for objects (Figure 5d) or even walking to the TV and back (Figure 5h).


**Figure 7. Word clouds generated from our participants' word selections from the Microsoft Reaction Cards [5] to describe TV audience silhouettes. Left: word cloud generated from all participants' words ($N$=249). Right: word cloud generated from participants' selections of top-5 most relevant words that describe audience silhouettes ($N$=60). Note the high frequency of positive words, such as *creative*, *fun*, *friendly*, *entertaining*, *connected*, and *collaborative*. NOTE: word clouds were generated with the on-line tool available at `http://www.wordle.net`.**

### Perceived experience and self-reported feedback

Participants rated their experience using 5-point Likert scales; see Figure 8. Overall, participants agreed that audience silhouettes made them feel connected to the remote persons (median rating 4 - *agree*), connectedness which they felt to improve their watching experience (median rating 4) for a TV application they perceived as useful (median rating 4). Also, participants considered that they were able to understand well the body movements of the on-screen silhouettes (median rating 4), which did not cause discomfort during television watching (median rating 2, *i.e., disagree* with the discomfort statement). Mann-Whitney $U$ tests did not detect any significant effect (at $p$=.05) of GENDER on any of these self-reported measures.


**Figure 8. Median values ($N$=15) for participants' self-reported experience collected with 5-point Likert scales (1 to 5): *strongly disagree*, *disagree*, *neither agree nor disagree*, *agree*, and *strongly agree*.**

We found significant positive correlations between perceived USEFULNESS and SOCIAL-EXPERIENCE (Spearman's $\rho_{(N=15)}$=.637, $p$=.05), between perceived DISTRACTEDNESS and SOCIAL-DISCOMFORT ($\rho_{(N=15)}$=.662, $p$=.01), as well as between perceived CONNECTEDNESS and KINESICS ($\rho_{(N=15)}$=.720, $p$=.01). These results show that participants that rated audience silhouettes as more useful, also felt that their television watching experience was more enjoyable, while participants that found silhouettes distracting also reported discomfort while watching TV with remote viewers. The degree of perceived connectedness to other viewers represented as silhouettes was higher when participants felt they understood the meaning of silhouettes' body movements (Spearman's $\rho_{(N=15)}$=.720, $p$=.01). We also found significant negative correlations between perceived USEFULNESS and DISTRACTEDNESS ($\rho_{(N=15)}$=−.641, $p$=.01) and between USEFULNESS and perceived KINESICS ($\rho_{(N=15)}$=−.534, $p$=.05), the latter suggesting that less understanding of the silhouettes' body movements may have caused a low perceived usefulness of the audience silhouettes concept.

Next to collecting participants' degree of agreement with Likert scale statements, we also ran two usability tests for the audience silhouette concept by employing the System Usability Score tool [9] and the Microsoft Reaction Cards [5].

The average SUS usability score computed from participants' self-reported answers to all the 10 questions of the test [9] was 68.3 (SD=8.3, $CI_{95\%}$=[63.8, 72.9]). Note that SUS scores range in [0, 100], with 100 representing a perfect usability result. Based on previous research employing SUS scores [2,3], our result is slightly above average, near the *good* threshold (that corresponds to SUS=70) in terms of the 7-point adjective ratings scale [2] (p. 121), and it falls within the *high acceptability* range proposed by Bangor *et al.* [3]. Men generally rated usability of audience silhouettes higher than women (70.0 versus 66.4), yet we found no statistically effect of GENDER on SUS ($U$=19.5, $Z$=−1.004, $n.s.$ at $p$=.05)

Our participants used an average of 16.6 words (SD=7.3) from the Microsoft Reaction Cards [5] to describe their experience with audience silhouettes. Figure 7 shows two word clouds generated from all participants' word selections ($N$=249 words, Figure 7, left) as well as from their top-5 most relevant words ($N$=60, Figure 7, right). Note the high frequency of positive words, such as *creative* (14, almost all participants considered audience silhouettes to be creative), *fun* (12), *friendly* (11), *entertaining* (9), *connected* (8), *innovative* (8), *attractive* (8), and *collaborative* (8). Women used in average more words than men to describe audience silhouettes (19.7, SD=9.2 versus 13.9, SD=4.1), however we did not detect any significant effect of GENDER, as showed by a Mann-Whitney $U$ test ($U$=16.000, $Z$=−1.392, $n.s.$).

### Open-ended feedback and comments

Next to objective and subjective measures, we also collected participants' open-ended comments about audience silhouettes. By analyzing those comments, we were able to identify several common perceptions, such as (i) audience silhouettes can help

viewers interact and communicate with each other using non-verbal behavior, (ii) they foster connectedness between remote persons, (iii) they may be used as indicators for TV content popularity (*i.e.*, what to watch?), and (iv) audience silhouettes may also interfere with one's privacy.

For example, the communication potential of audience silhouettes was repeatedly remarked by participants, *e.g.*, *"viewers can interact using non-verbal communication"* ($P_1$), silhouettes *"transmit emotions"* ($P_3$), *"help compare one's reaction to others"* ($P_4$), *"can replace verbal communication during television watching"* ($P_4$), they *"make television more interactive"* ($P_6$), *"help communication"* ($P_7$), and *"help interaction between viewers"* ($P_{14}$). Participants also remarked the capability of audience silhouettes to deliver the feeling of connectedness between remote individuals, *e.g.*, silhouettes *"reduce the feeling of loneliness when watching TV"* ($P_1$), *"it is comforting to know that my friends are all right"* ($P_8$), *"less loneliness when watching TV"* ($P_8$), *"I would feel more close to dear ones, making sure the other person is all right"* ($P_{10}$), *"watching movies becomes a social experience"* ($P_{12}$), *"helps with the feeling of loneliness while watching TV"* ($P_{14}$), silhouettes are a *"form of socialization"* ($P_9$) and a *"tool for socialization"* ($P_{11}$). Several participants suggested the use of audience silhouettes as indicators for the quality of TV content, *i.e.*, more silhouettes means more of their friends enjoying the show: *"the number of silhouettes are a recommender for that show"* ($P_7$), *"I can discover what others are watching, what are their preferences"* ($P_{13}$), *"useful feedback for evaluating the quality of a show"* ($P_{13}$). Some participants had privacy concerns in the case of a security breach in the system.

Eleven (11) of our participants (73%) said they would like to have the audience silhouettes application running on their home TVs, and 13 of them (87%) said they would recommend the audience silhouettes system to others.

## CONCLUSION AND FUTURE WORK

We introduced in this work *audience silhouettes* as a practical technique to convey *peripheral awareness* of remote viewers and to leverage *kinesics* as a non-intrusive communication channel for viewers during television watching. Overall, our participants rated the concept favorably, a finding that we were able to verify with multiple usability metrics. Furthermore, we characterized participants' body movement responses in relation to the on-screen audience silhouettes, and we introduced a visualization technique to serve for future explorations in this line of work. The data that we collected enable us to believe that audience silhouettes can provide a simple and effective channel for presence and non-verbal communication over distance in the context of social television watching.

Future work will address *in-situ* studies (for which we expect more body movement reactions), solving technical issues for transmitting many silhouettes over the network without compromising video synchronization [22], and visualization enhancements (*e.g.*, color, depth granularity, etc.). Also, exploration of more body movement measures (*e.g.*, kinematic features relying on speed and acceleration, which we did not explore in this work) will likely reveal more findings about how viewers engage with interactive television content. Predicting viewers' body behavior and engagement with television content with workable models [6] would be very valuable for designers of such iTV systems. Interactive TV systems will probably benefit of combining audience silhouettes with whole-body gesture recognition [41] that will offer users the opportunity for more control, from passive engagement (*i.e.*, using the silhouette only, as in this work) to gesture commands that invoke specific functions for iTV [42,43,46]. The effect of distributed visual attention on peripheral awareness of various audiences for multi-screen TV [44,45] is also an interesting research direction.

We believe that this work on audience silhouettes has barely scratched the opportunity of employing real-time kinesics for social TV watching, and we are eager to see how the community will employ our concept and techniques to design enriched social interactive television experiences for viewers.

## ACKNOWLEDGMENTS

This work was supported by the liFe-StaGE project 740/2014, "Multimodal Feedback for Supporting Gestural Interaction in Smart Environments", co-funded by UEFISCDI & OeAD.

## REFERENCES

1. Analysis Manson Limited. The Connected Consumer Survey 2013: TV and video **http://www.analysysmason.com/Research/Content/Reports/Connected-Consumer-TV-May2013-RDMB0/samples-TOC/** (last accessed march 2015).
2. Bangor, A., Kortum, P., and Miller, J. Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of Usability Studies 4*, 3 (2009), 114–123.
3. Bangor, A., Kortum, P. T., and Miller, J. T. An empirical evaluation of the system usability scale. *Int. Journal of Human-Computer Interaction 24*, 6 (2008), 574–594.
4. Basapur, S., Mandalia, H., Chaysinh, S., Lee, Y., Venkitaraman, N., and Metcalf, C. FANFEEDS: Evaluation of socially generated information feed on second screen as a TV show companion. In *Proc. of EuroITV '12*, ACM (New York, NY, USA, 2012), 87–96.
5. Benedek, J., and Miner, T. Measuring desirability: New methods for evaluating desirability in a usability lab setting. In *Proc. of the Usability Professionals Assoc. Conf.*, 2002 **http://www.microsoft.com/usability/uepostings/desirabilitytoolkit.doc**.
6. Bianchi-Berthouze, N. Understanding the role of body movement in player engagement. *Human Computer Interaction 28*, 1 (2013), 40–75.
7. Birdwhistell, R. Introduction to kinesics: An annotation system for analysis of body motion and gesture. Washington, DC: Foreign Service Institute, 1952.
8. Bobick, A. F., and Davis, J. W. The recognition of human movement using temporal templates. *IEEE TPAMI 23*, 3 (Mar. 2001), 257–267.
9. Brooke, J. SUS: A quick and dirty usability scale. In *Usability evaluation in industry*. Taylor & Francis, 1996.
10. Cesar, P., Bulterman, D. C., and Jansen, A. J. Usages of the secondary screen in an interactive television

environment: Control, enrich, share, and transfer television content. In *Proc. EuroITV'08* (2008), 168–177.

11. Cesar, P., Chorianopoulos, K., and Jensen, J. F. Social television and user interaction. *Computers in Entertainment 6*, 1 (May 2008), 4:1–4:10.
12. Chartrand, T., and Bargh, J. The chameleon effect: The perception-behavior link and social interaction. *J. of Personality and Social Psychology 76*, 6 (1999), 893–910.
13. Chorianopoulos, K. Content-enriched communication supporting the social uses of TV. *The Journal of The Communications Network 6*, 1 (2007), 23–30.
14. Coppens, T., Trappeniers, L., and Godon, M. AmigoTV: towards a social TV experience. In *EuroITV '04* (2004).
15. Courtois, C., and D'heer, E. Second screen applications and tablet users: Constellation, awareness, experience, and interest. In *Proc. of EuroITV '12* (2012), 153–156.
16. Dezfuli, N., Khalilbeigi, M., Huber, J., Müller, F., and Mühlhäuser, M. PalmRC: Imaginary palm-based remote control for eyes-free television interaction. In *Proc. of EuroITV '12*, ACM (New York, NY, USA, 2012), 27–34.
17. Doughty, M., Rowland, D., and Lawson, S. Who is on your sofa?: TV audience communities and second screening social networks. In *EuroITV '12* (2012), 79–86.
18. Feyereisen, P., and de Lannoy, J. *Gestures and Speech: Psychological Investigations*. Cambridge University Press, New York, 1991.
19. Geerts, D. Comparing voice chat and text chat in a communication tool for interactive television. In *Proc. of NordiCHI '06*, ACM (New York, USA, 2006), 461–464.
20. Geerts, D., Cesar, P., and Bulterman, D. The implications of program genres for the design of social television systems. In *Proc. of UXTV '08* (2008), 71–80.
21. Geerts, D., and De Grooff, D. Supporting the social uses of television: Sociability heuristics for social TV. In *Proc. of CHI '09*, ACM (New York, NY, USA, 2009), 595–604.
22. Geerts, D., Vaishnavi, I., Mekuria, R., van Deventer, O., and Cesar, P. Are we in sync?: Synchronization requirements for watching online video together. In *Proc. of CHI '11*, ACM (New York, NY, USA, 2011), 311–314.
23. Graham, J. A., and Argyle, M. A cross-cultural study of the communication of extra-verbal meaning by gestures. *Int. Journal of Psychology 10* (1975), 57–67.
24. Hall, J. A. *Nonverbal Sex Differences: Communication Accuracy and Expressive Style Paperback*. Johns Hopkins University Press, 1990.
25. Harboe, G., Massey, N., Metcalf, C., Wheatley, D., and Romano, G. The uses of social television. *Computers in Entertainment 6*, 1 (May 2008), 8:1–8:15.
26. Harboe, G., Metcalf, C. J., Bentley, F., Tullio, J., Massey, N., and Romano, G. Ambient social TV: Drawing people into a shared experience. In *CHI '08* (2008), 1–10.
27. Hawkins, R. P., Pingree, S., Hitchon, J., Radler, B., Gorham, B. W., Kahlor, L., Gilligan, E., Serlin, R. C., Schmidt, T., Kannaovakun, P., and Kolbeins, G. H. What produces television attention and attention style? *Human Communication Research 31*, 1 (2005), 162–187.
28. Huang, E. M., Harboe, G., Tullio, J., Novak, A., Massey, N., Metcalf, C. J., and Romano, G. Of social television comes home: A field study of communication choices and practices in TV-based text and voice chat. In *Proc. of CHI '09*, ACM (New York, NY, USA, 2009), 585–594.
29. Jones, B., Sodhi, R., Murdock, M., Mehra, R., Benko, H., Wilson, A., Ofek, E., MacIntyre, B., Raghuvanshi, N., and Shapira, L. Roomalive: Magical experiences enabled by scalable, adaptive projector-camera units. In *Proc. of UIST '14*, ACM (New York, NY, USA, 2014), 637–644.
30. Jones, B. R., Benko, H., Ofek, E., and Wilson, A. D. Illumiroom: Peripheral projected illusions for interactive experiences. In *Proc. of CHI '13* (2013), 869–878.
31. Kelley, J. F. An iterative design methodology for user-friendly natural language office information applications. *ACM Trans. on Inf. Sys. 2*, 1 (1984), 26–41.
32. Kendon, A. Do gestures communicate? A review. *Research on Language and Social Interaction 27* (1994), 175–200.
33. Kleinsmith, A., and Bianchi-Berthouze, N. Affective body expression perception and recognition: A survey. *IEEE Trans. Affect. Comput. 4*, 1 (Jan. 2013), 15–33.
34. McNeill, D. *Hand and Mind: What Gesture Reveals about Thought*. University Chicago Press, 1992.
35. Morris, M. R. Web on the wall: Insights from a multimodal interaction elicitation study. In *Proc. of ITS '12*, ACM (New York, NY, USA, 2012), 95–104.
36. Nathan, M., Harrison, C., Yarosh, S., Terveen, L., Stead, L., and Amento, B. CollaboraTV: Making television viewing social again. In *Proc. UXTV '08* (2008), 85–94.
37. Oehlberg, L., Ducheneaut, N., Thornton, J. D., Moore, R. J., and Nickell, E. Social TV: Designing for distributed, sociable television viewing. In *Proc. of EuroITV '06* (2006), 251–259.
38. Ren, G., and O'Neill, E. Freehand gestural text entry for interactive TV. In *Proc. of EuroITV '13* (2013), 121–130.
39. Schirra, S., Sun, H., and Bentley, F. Together alone: Motivations for live-tweeting a television series. In *Proc. of CHI '14*, ACM (New York, USA, 2014), 2441–2450.
40. Shamma, D. A., Bastea-Forte, M., Joubert, N., and Liu, Y. Enhancing online personal connections through the synchronized sharing of online video. In *Proc. of CHI EA '08*, ACM (New York, NY, USA, 2008), 2931–2936.
41. Vatavu, R.-D. Nomadic gestures: A technique for reusing gesture commands for frequent ambient interactions. *J. Ambient Intell. Smart Environ. 4*, 2 (Apr. 2012), 79–93.
42. Vatavu, R.-D. User-defined gestures for free-hand TV control. In *Proc. of EuroiTV '12* (2012), 45–48.
43. Vatavu, R.-D. A comparative study of user-defined handheld vs. freehand gestures for home entertainment environments. *Journal of Ambient Intelligence and Smart Environments 5*, 2 (2013), 187–211.
44. Vatavu, R.-D. There's a world outside your TV: Exploring interactions beyond the physical TV screen. In *EuroITV '13*, ACM (New York, NY, USA, 2013), 143–152.
45. Vatavu, R.-D., and Mancas, M. Visual attention measures for multi-screen tv. In *Proc. of TVX '14*, ACM (New York, NY, USA, 2014), 111–118.
46. Vatavu, R.-D., and Zaiti, I.-A. Leap gestures for TV: Insights from an elicitation study. In *Proc. of TVX '14*, ACM (New York, NY, USA, 2014), 131–138.

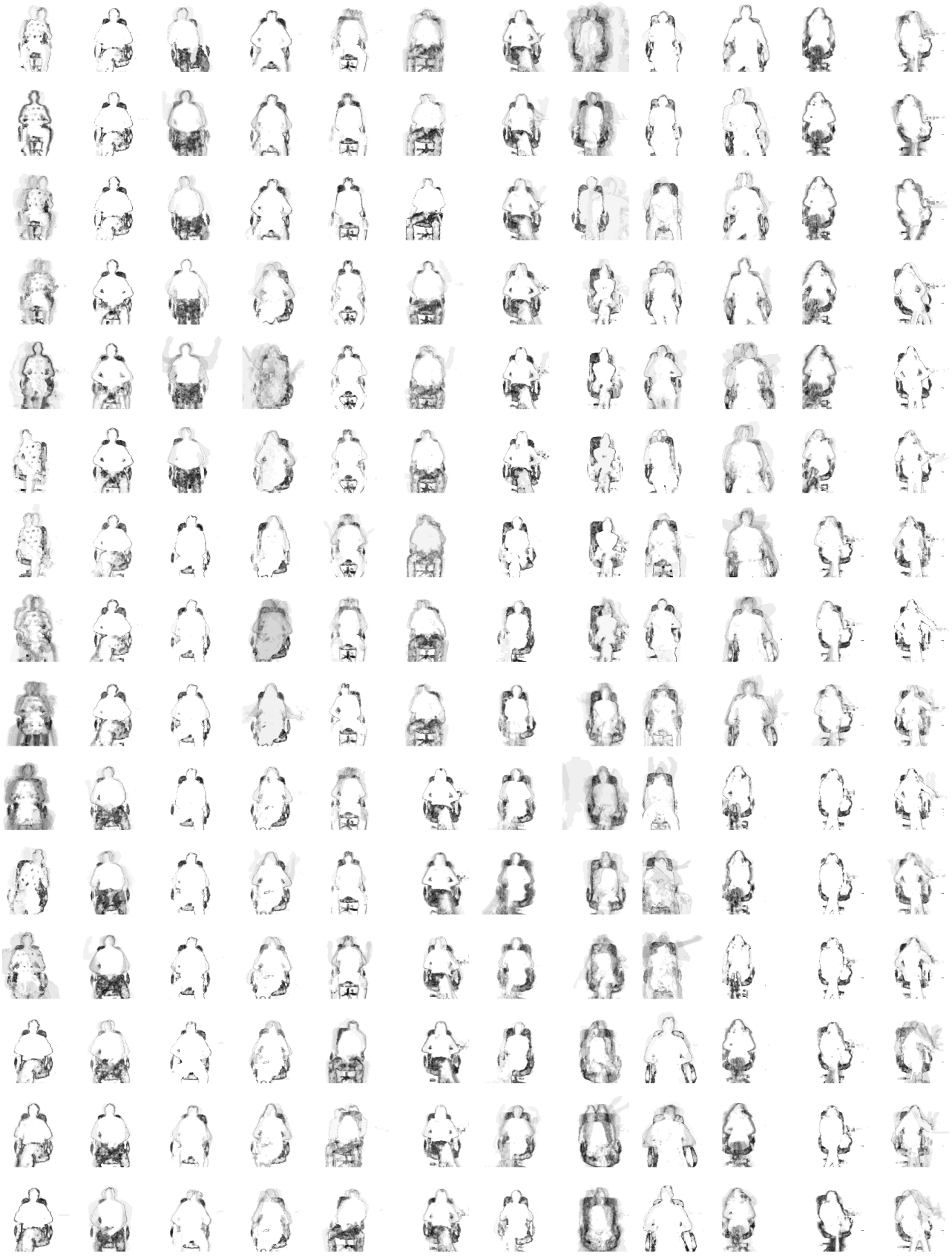

**Figure 9. Motion-Amount Images for all the 180 experimental conditions (=15 participants $\times$ 4 AUDIENCE-TYPES $\times$ 3 GENRES).**